Army Data Services Layer (ADSL) − Data Mediation Providing Data Interoperability and Understanding in a SOA Environment

Michelle Dirner
Army Net-Centric Data Strategy (ANCDS) Center of Excellence (CoE)
Data Service Team Lead
RDECOM CERDEC SED in support of SEC
Fort Monmouth, NJ
michelle.dirner@us.army.mil

Michael Mansell
Army Net-Centric Data Strategy (ANCDS) Center of Excellence (CoE)
Data Mediation Service Project Lead
RDECOM CERDEC SED in support of SEC
Fort Monmouth, NJ
michael.mansell2@us.army.mil

- **DoD Data Strategy Vision**

- **Army Data Strategy**

- **Army Data Strategy Center of Excellence**

- **Data and SOA Together**

- **What is SOA?**

- **What is ADSL?**

- **Data Mediation Service**

- **ADSL and Army SOA Foundation (SOAF-A)**

Thanks to advancements in and the ubiquity of computing technology, there is more data available now than ever before.

However, the ability to process and understand the data has not scaled as quickly as the ability to generate and collect data, thus leading to an environment with the following characteristics:

- Large amounts of currently available data.

- A high rate of new data being generated.

- Contradictions in available data.

- An imbalance of relevant data compared to data found.

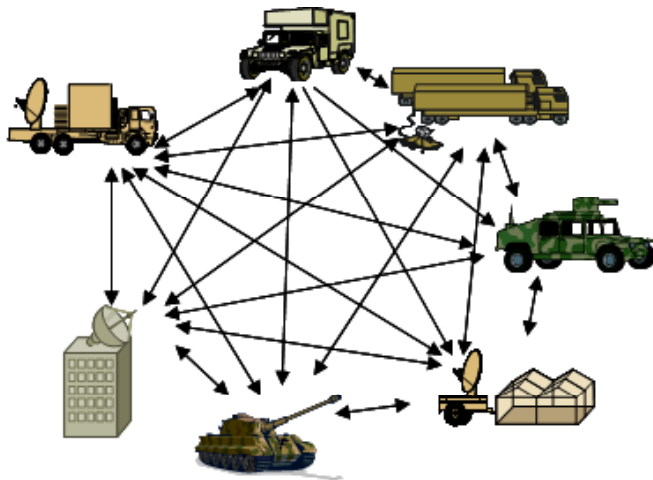- Inefficient methods for comparing and processing different kinds of data.

As a result, decision-makers are inundated with large amounts of data resulting in more time spent analyzing data rather than timely action.

Current

- ➢ Pre-determined "point to point" connections between systems and applications on disparate networks

- ➢ Producer "pushes" information to pre-defined consumers

Future

- ➢ Authorized known and unanticipated consumers access data they need regardless of who produced the information

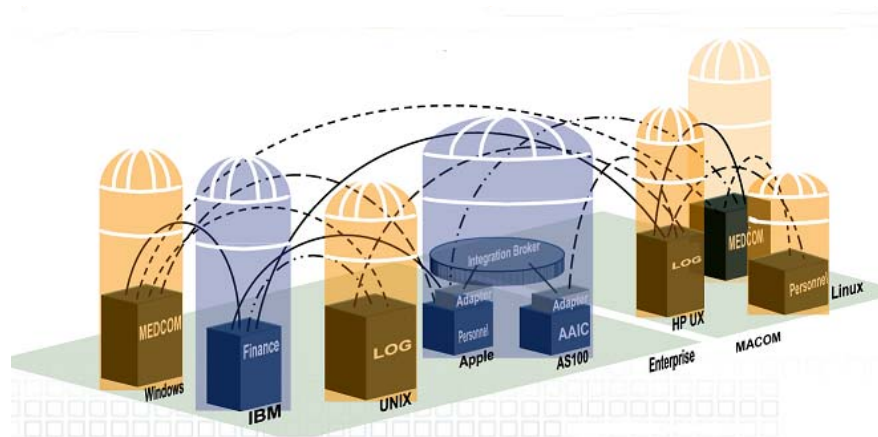- ➢ Systems and applications are web-enabled to expose their information

Information not easily shared

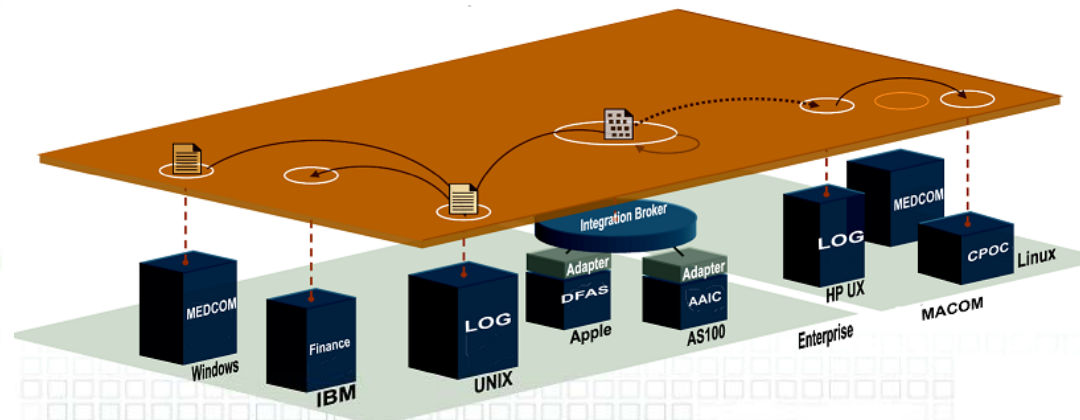
Information Ubiquitous on the Net

4

Current

- Few authoritative sources identified
- Complexity of inter-organizational collaboration
- Need for data conversion impacts timeliness of military decision-making process
- Cultural and technical impediments to data sharing

Future

- Federated Enterprise
- Platform independent data and application reuse
- Easily defined and updated business processes
- Standardized interface specifications and authoritative data sources
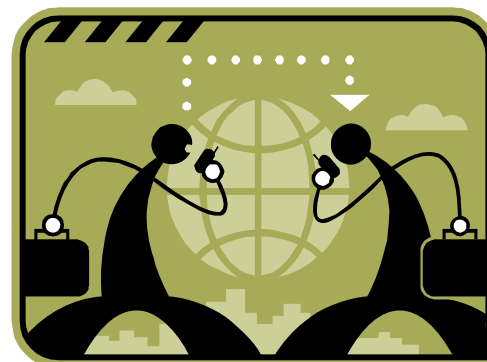
Net-Centricity Goals

1. Visible – <u>Who</u> has <u>what</u> data available?

2. Accessible – <u>Where</u> is this data and in what format?

3. Understandable – <u>What</u> does this data mean?

4. Trusted – Is this data <u>trustworthy</u>, accurate, and authoritative?

5. Interoperable – Can my application <u>use</u> the data?

6. Responsive to users needs – Is the data applicable and <u>timely</u>?

7. Institutionalized – <u>What</u> and <u>who</u> governs the definition, lifecycle, and use of this data?
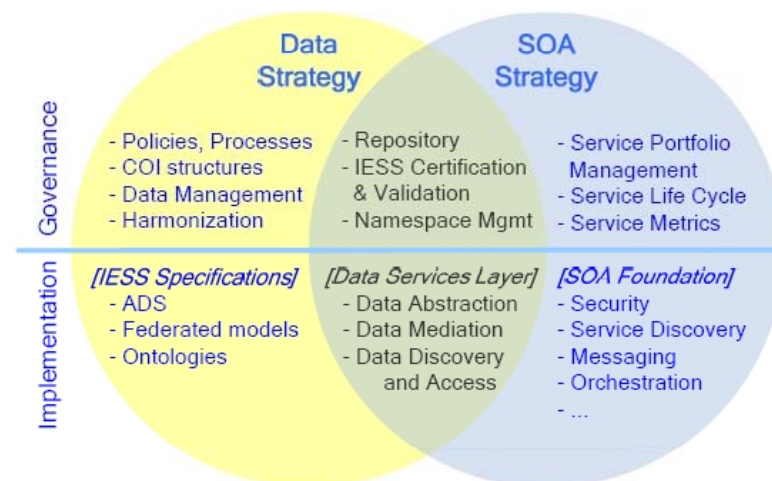
- **The Army Net-Centric Data Strategy (ANCDS) Center of Excellence (CoE) was established to facilitate the execution of the Army Data Strategy while providing users with overarching data products and services to promote interoperability and faster access, retrieval, analysis and utilization of data.**

- **The CoE provides uniform expertise to enable the migration to a net-centric environment via the establishment, production, and implementation of Communities of Interest (COIs) and their products.**

- **The DS CoE supports the lifecycle from conception to the field by providing the services for data requirements generation, implementation, integration, and validation in support of the Army Data Strategy.**
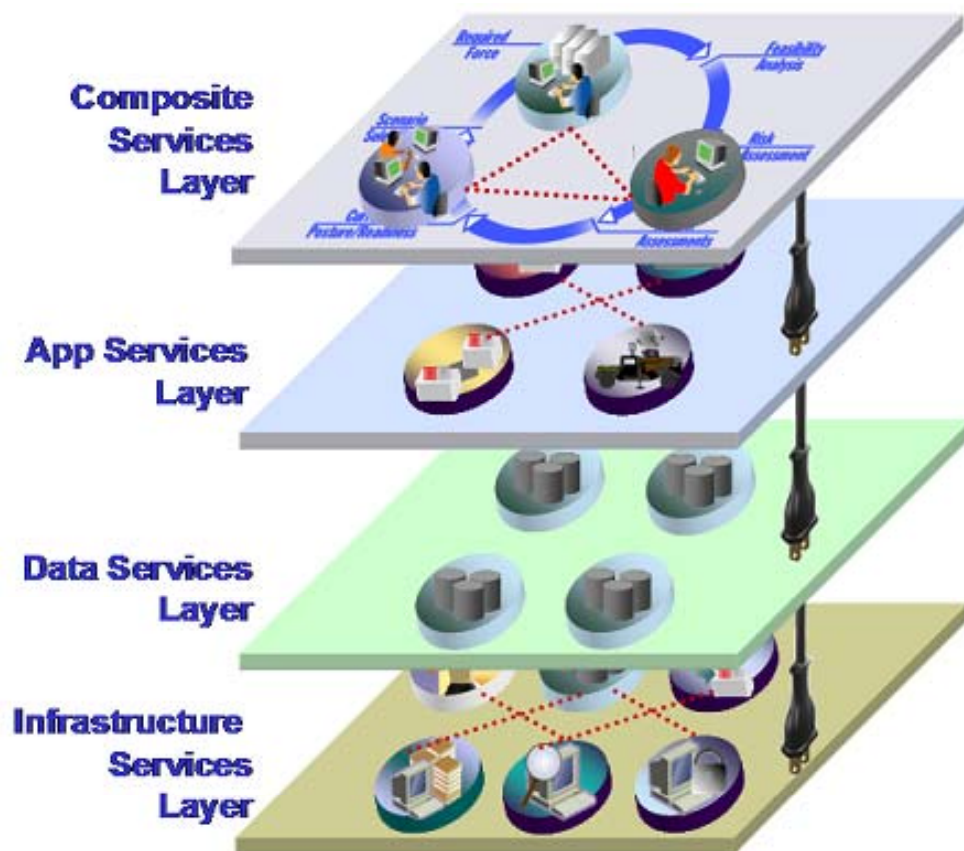
➢ **Data Strategy** prescribes a strategy for identifying the data to be shared, where that data should be coming from (authoritative sources) and standard representations for sharing that data

➢ **SOA Strategy** creates the services environment that addresses the details for a common framework and mechanisms for sharing data across heterogeneous environments in a seamless manner

Data Strategy

- Policies, Processes
- COI structures
- Data Management
- Harmonization

- Repository
- IESS Certification & Validation
- Namespace Mgmt

SOA Strategy

- Service Portfolio Management
- Service Life Cycle
- Service Metrics

Governance

Implementation

[IESS Specifications]
- ADS
- Federated models
- Ontologies

[Data Services Layer]
- Data Abstraction
- Data Mediation
- Data Discovery and Access

[SOA Foundation]
- Security
- Service Discovery
- Messaging
- Orchestration
- ...

➢ *ADSL is common ground between SOA Strategy and Data Strategy that enables sharing and management of data that is distributed across the enterprise*

➢ *ADSL is not a single physical capability; instead it is a virtual data tier for the Army enterprise – accessible and available on the LandWarNet*

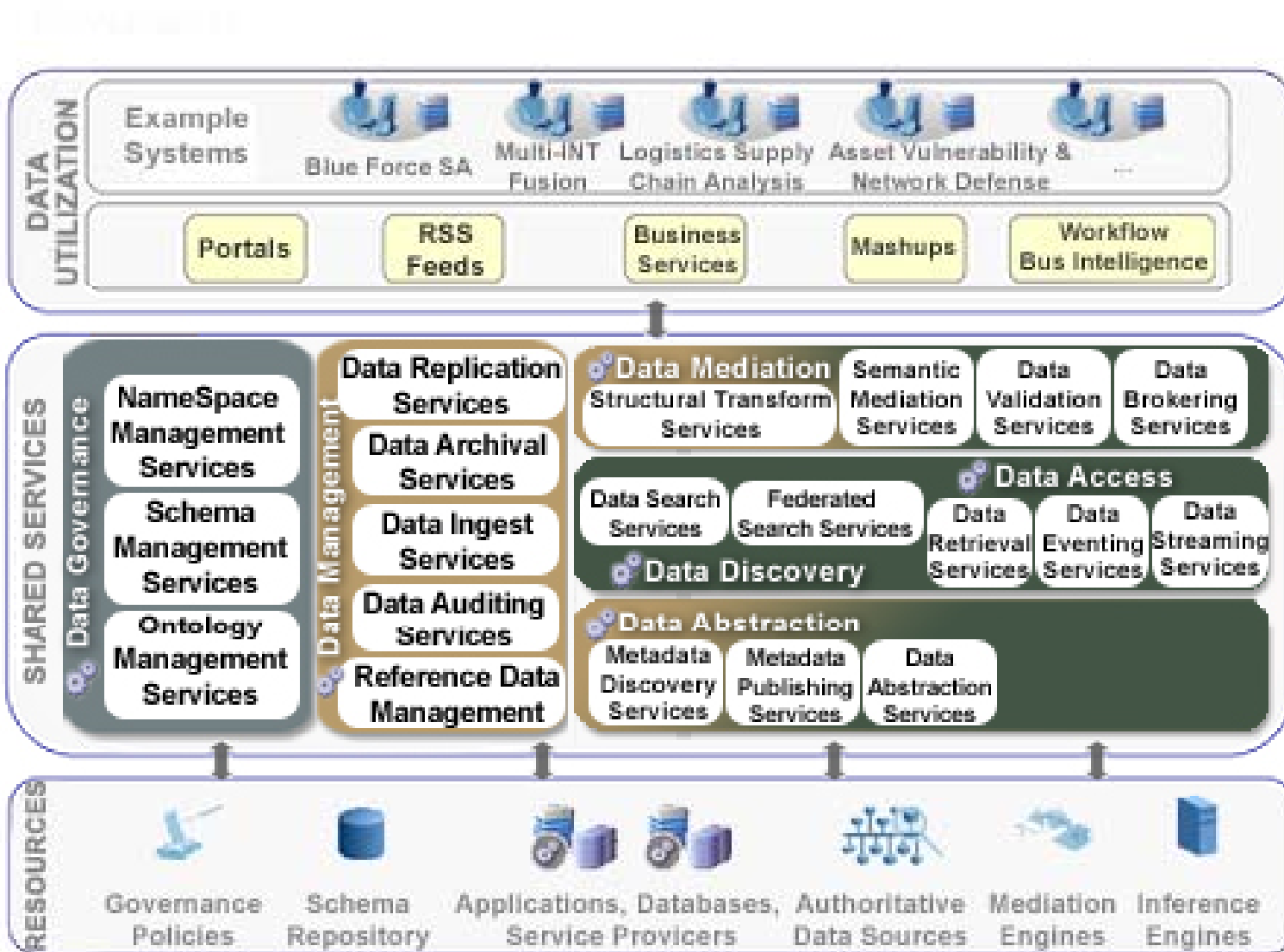*Two Complementary Strategies to Achieve Net Centricity*

- SOA is a layered architecture and the Data Services Layer is one of these layers

- Army Data Services Layer is closest to the raw data sources in this architecture

- Architecturally speaking, data services combine to form a middle layer of reusable services

- Data services are a form of web service, optimized to meet real-time data integration demands of SOA

9

UNCLASSIFIED

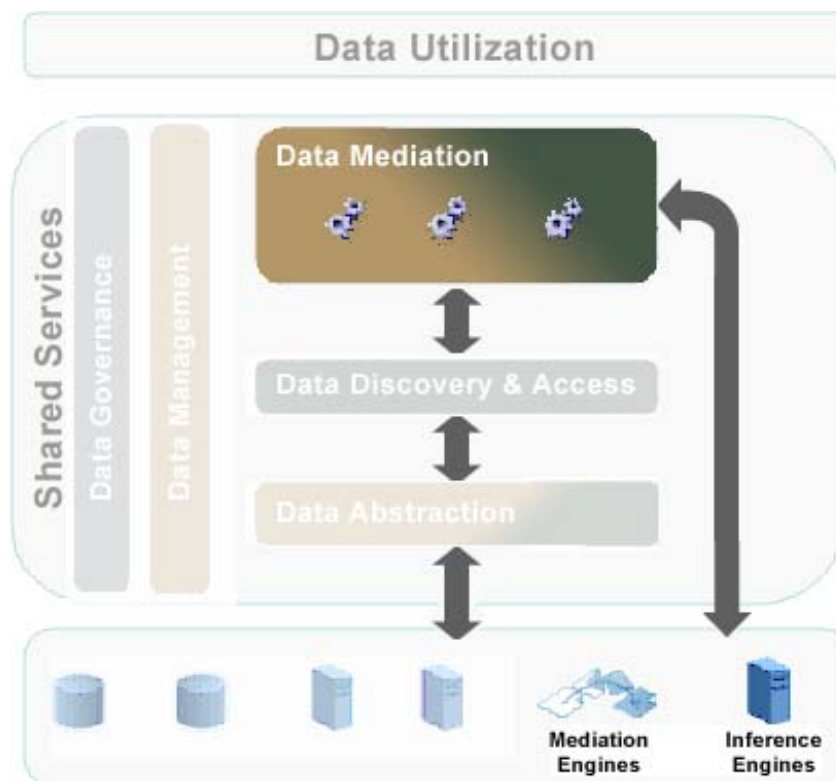The ADSL consists of the following families of data services:

- Data Utilization: Consists of end-to-end composable applications that are built upon the other "atomic" data services. These applications are user-oriented and directly enable warfighters and decision-makers to use data inside and outside of the Army to satisfy mission needs.

- Data Mediation: Bridging the gap among different data formats, vocabularies, and semantics, making data understandable and usable to the consumers who are otherwise unable to make use of the data.

- Data Abstraction: Capturing and managing the metadata, at both structural and semantic levels. Such metadata is made available across the Enterprise, allowing for data to be visible and discoverable.

- Data Access: Exposing interfaces to search, retrieve, and manipulate data.

- Data Management: Providing the persistence and management of data "at-rest".

- Data is available, in a large enterprise such as the Army, in many different formats presenting semantic and syntactic challenges

- Data Mediation bridges the gap among disparate data types, vocabularies, and semantics

- Data Mediation is a family of services that enables a consumer to understand data from other sources to produce a coherent set of information, by making use of semantic mapping, structural and syntactic transforms, inferencing and validation

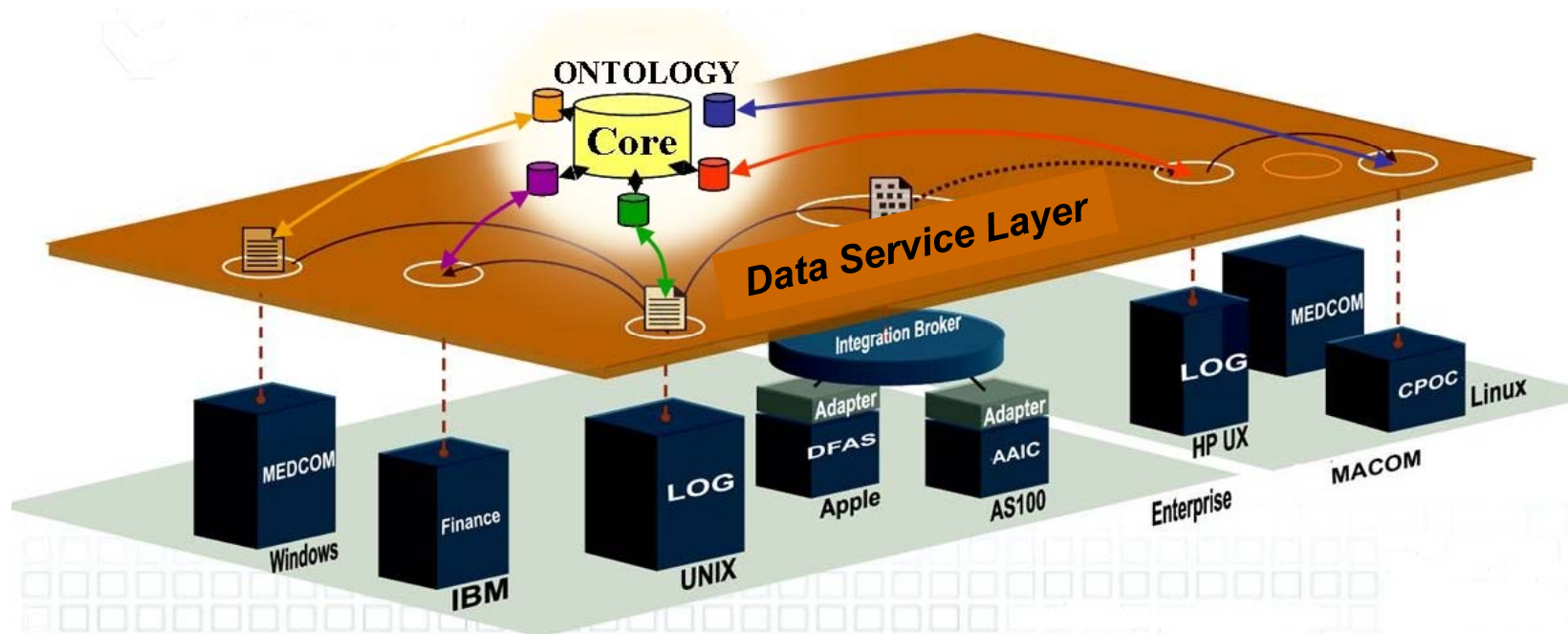**Data Mediation** facilitates usability through:

- **Structural and Syntactic Transformation**
  - Applies XSL Transform to XML messages

- **Semantic Mediation**
  - Builds on structural transformation by automating XSLT creation process
  - Inference Engines – are used to reason about an Ontology, examining concepts and their relationships and logically drawing conclusions to create concept mappings
  - Mediation Engines – use concept mappings provided by the inference engine and apply them to an XSLT structure (DMS specific)

- **Data Validation**
  - Ensure that accurate data mediation has taken place
  - Can also validate business rules

- **Data Brokering**
  - Automate / facilitate interaction between other services and resources

13

- Semantics provide the foundation of the Data Mediation Service operation

- How the Data Mediation Service utilizes semantics:

  – Ontology in SOA

  – What Is an Ontology?

  – Web Ontology Language (OWL)

  – Semantic Mapping

  – DMS Operation and Design

14

Ontology technology serves a foundational part of the Data Service Layer by facilitating the following goals:

- Reduction or elimination (where possible) of n-squared problem
- Increase in data quality and decrease in errors and inconsistency
- Increase in flexibility, reuse and efficiency
- Decrease in the amount of maintenance required
- **Increase in automation of data mediation**

An ontology is an abstract model of a domain that consists of concepts, attributes/properties, inter-concept relationships and logic

Ontology components:
- concepts
- properties and attributes
- inter-concept relationships
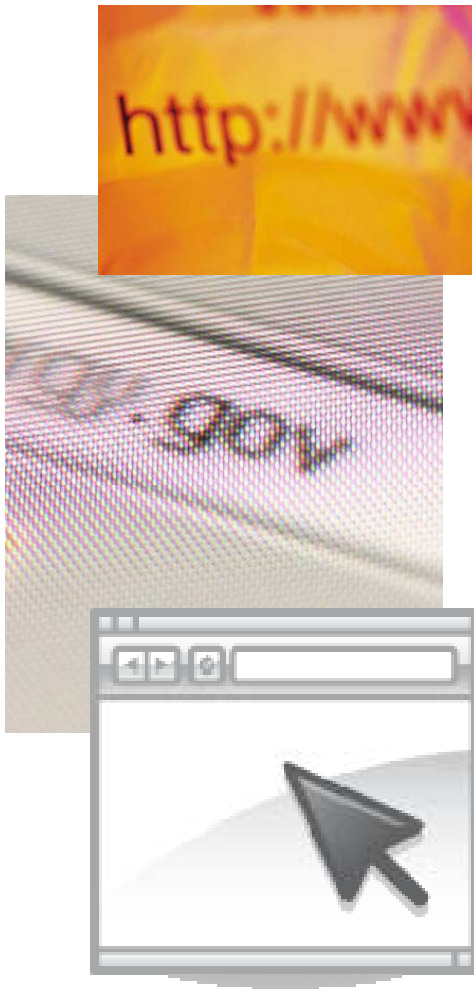- individuals
- business rules/domain logic

An ontology defines:
- a common vocabulary
- a shared understanding

An ontology is used to provide a shared common understanding of a data domain. It is expressed in machine-readable code, and captures data domain logic/business rules as data within it. When combined with inference engines, an ontology is used to perform inferencing (reasoning).
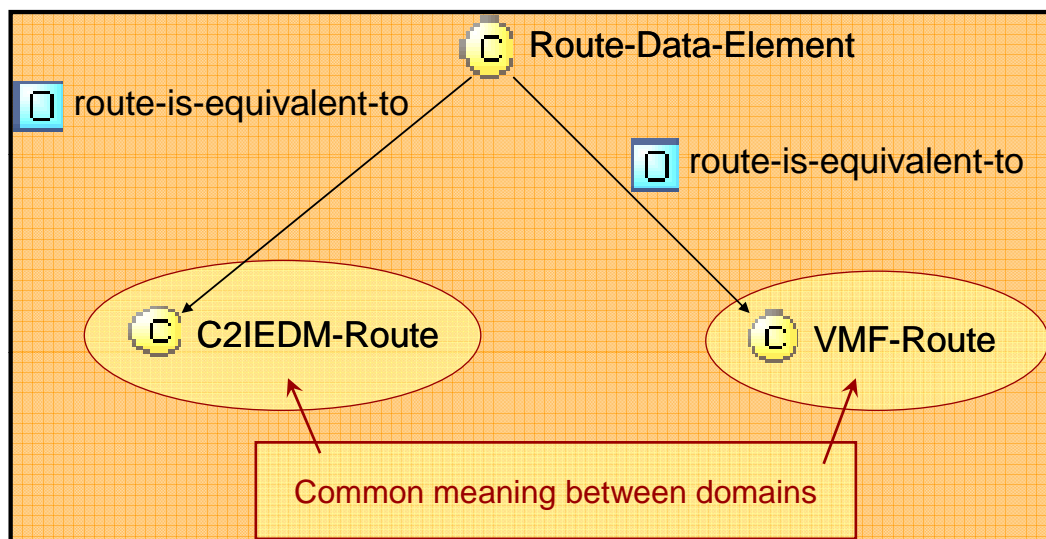
16

Web Ontology Language (OWL)

- **The** World Wide Web Consortium **(W3C) is the main international standards organization for the World Wide Web.**

- **The** Web Ontology Language **(OWL)**

  - **a language for defining and instantiating Web ontologies**

  - **a W3C recommendation**

  - **designed to be read by computer applications**

  - **based on** eXtensible Markup Language **(XML)**

  - **acknowledged as a major technology for the future implementation of the Semantic Web**

- OWL Ontology Development Tools

  - TopBraid Composer **is a professional development environment specifically used for the OWL and is the ontology tool chosen for ANCDS CoE**

  - Protégé 2000 **is a free, open source ontology editor**

The Data Mediation Service is designed for OWL-DL.

Route-Data-Element

route-is-equivalent-to

route-is-equivalent-to

C2IEDM-Route
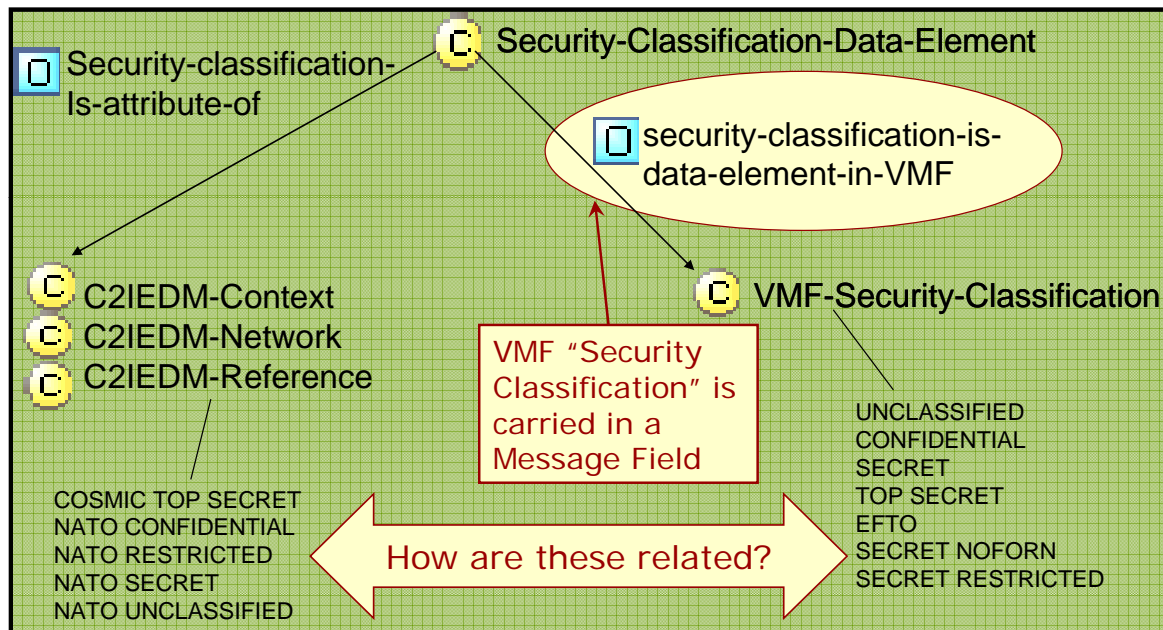
VMF-Route

Common meaning between domains

Mapping Example 1:

The concept of "route" as defined in the C2IEDM and VMF.

Mapping Example 2:

The concept of "Security-Classification" as defined in the C2IEDM and VMF.

Security-Classification-Data-Element

Security-classification-Is-attribute-of

security-classification-is-data-element-in-VMF

C2IEDM-Context
C2IEDM-Network
C2IEDM-Reference

VMF-Security-Classification

VMF "Security Classification" is carried in a Message Field

COSMIC TOP SECRET
NATO CONFIDENTIAL
NATO RESTRICTED
NATO SECRET
NATO UNCLASSIFIED

How are these related?

UNCLASSIFIED
CONFIDENTIAL
SECRET
TOP SECRET
EFTO
SECRET NOFORN
SECRET RESTRICTED

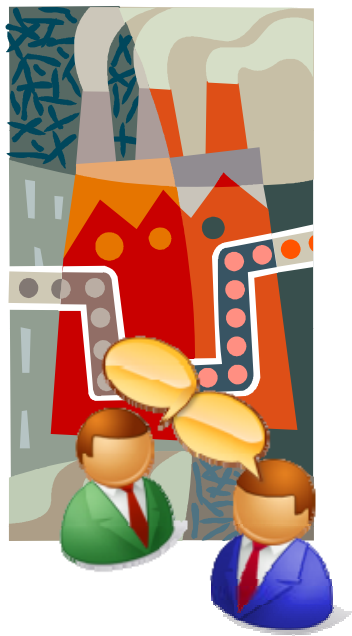- **The DMS requires two pieces of information to start.**
  - **The source identifier (namespace/URI/URN)**
  - **The target identifier (namespace/URI/URN)**

- **Once these are provided the DMS queries the ADSL Repository for eXtensible Markup Language (XML) Schemas and Web Ontology Language (OWL) Ontologies related to the provided source and target identifiers.**

- **The DMS uses the semantics and business rules inherent in OWL Ontologies and processes them in an Inference Engine to create concept mappings.**
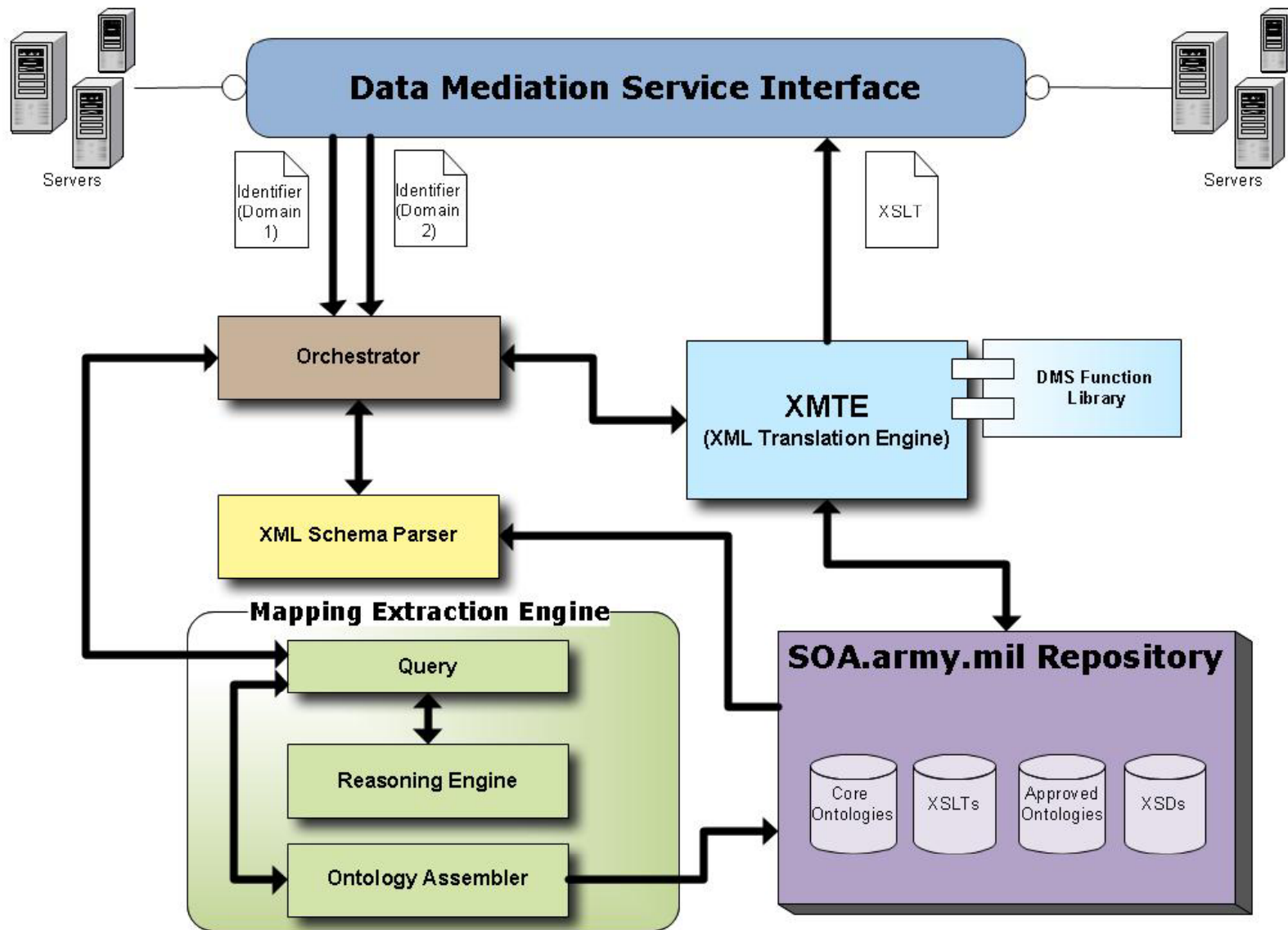
- **The DMS then uses the strengths of XML to provide structural and syntactic information and combines this with concept mappings created by the Inference Engine to create accurate and reusable XSL Transforms (XSLTs) on demand.**

- **These XSLTs are provided to the calling service, application, or ESB to facilitate quick, accurate, semantically enabled data mediation.**
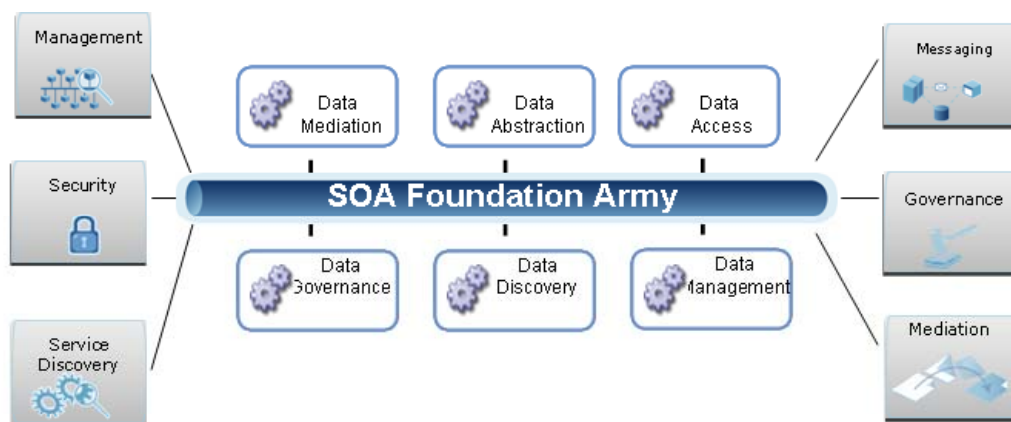
19

UNCLASSIFIED

- SOA Foundation defines a set of infrastructure services within Army, which are compliant and interoperable with NCES standards

  - Service Discovery

  - Security

  - Messaging

  - Governance

- A Reference Implementation of ADSL will be hosted on SOAF-A

For more information on the data services work being conducted by the Army Data Strategy Center of Excellence, please contact:

Michelle Dirner
ANCDS COE Data Service Team Lead
RDECOM CERDEC SED in support of SEC
Fort Monmouth, NJ 07703
michelle.dirner@us.army.mil
Office: (732) 532-3304

Michael Mansell
ANCDS COE Data Mediation Service Tech Lead
RDECOM CERDEC SED in support of SEC
Fort Monmouth, NJ 07703
michael.mansell2@us.army.mil
Office: (732) 532-2235